



2

AD-A231 304

Assessment of Differential Prediction by Race For the USNA Classes of 1986-1990

**Steven E. Devlin
Norman M. Abrahams**

DTIC
ELECT
JAN 28 1991
S B D

**Assessment of Differential Prediction by Race
For the USNA Classes of 1986-1990**

Steven E. Devlin
Navy Personnel Research and Development Center

Norman M. Abrahams
Personnel Decisions Research Institute
Minneapolis, MN 55414

Reviewed by
Jack E. Edwards

Approved by
Patricia J. Thomas

Released by
Jules I. Borack
Director, Personnel Systems Department

Approved for public release;
distribution is unlimited.

Navy Personnel Research and Development Center
San Diego, California 92152-6800

**THIS DOCUMENT CONTAINED
BLANK PAGES THAT HAVE
BEEN DELETED**

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE December 1990		3. REPORT TYPE AND DATE COVERED Tech. Note--Aug 1989-Feb 1990
4. TITLE AND SUBTITLE Assessment of Differential Prediction by Race for the USNA Classes of 1986-1990			5. FUNDING NUMBERS Program Element 0602233N Work Unit RM33M20.10	
6. AUTHOR(S) Steven E. Devlin and Norman M. Abrahams				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Navy Personnel Research and Development Center San Diego, California 92152-6800			8. PERFORMING ORGANIZATION REPORT NUMBER NPRDC-TN-91-2	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Technology Arlington, Virginia 22217-5000			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The USNA uses the Candidate Multiple (CM), a composite of several measures, to evaluate applicants. This composite seeks to predict many aspects of USNA performance. While the USNA strives for equal opportunity, the CM has not been evaluated for fairness for some time. Toward that end, the CM was evaluated using the widely accepted regression model of test fairness. This model effectively states that a test is biased for a subgroup if it consistently over- or underpredicts the performance for members of that group. Using this model, procedures were used to statistically assess whether a prediction system based on the CM is fair to Blacks when the system for predicting academic performance is developed on a combined sample of Blacks and Whites. In essence, these procedures are aimed at determining whether a separate regression equation (i.e., prediction system) for members of the minority group would provide a more accurate and equitable prediction of their performance. These analyses revealed a statistically significant difference between separate prediction systems. This difference, while statistically significant, increases prediction by only about three-fourths of one percentage point. Thus, the use of separate equations does not provide a meaningful improvement in the prediction of academic performance. Finally, since Black academic performance scores were generally overpredicted when the combined-group regression equation was used, a combined-group regression equation does not adversely impact upon the Black minority. It is therefore recommended that the USNA continue using a single prediction system.				
14. SUBJECT TERMS United States Naval Academy, selection, fairness, minorities, differential prediction			15. NUMBER OF PAGES 15	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UNLIMITED	

FOREWORD

This report assesses fairness in selection of Blacks applying for admission to the United States Naval Academy. An assessment of differential prediction by race addresses this concern while clarifying some of the issues involved. Various statistical techniques are employed to determine if the selection process is biased toward Blacks. The results may be of use to the Admissions Office, United States Naval Academy.

This effort was conducted under the sponsorship of the Office of Naval Technology (Code 222) within exploratory development (Program Element 0602233N, Project Task RM33M20.10 Selection Systems for Changing Organizations). This is the first report within this program element and work unit number.

JULES I. BORACK
Director, Personnel Systems Department

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



SUMMARY

Problem and Background

The United States Naval Academy (USNA) selection process uses a single selection composite for applicants, minority and majority. This composite, the Candidate Multiple (CM), combines information from various sources to help predict academic and other aspects of USNA performance. To provide equal opportunity for Blacks, it is especially important to ensure that a single prediction system (equation) is appropriate for predicting the performance of different racial groups.

A series of statistical procedures can be used to determine whether a single equation is equitable. Specifically, the need for separate prediction equations based on group membership is demonstrated when the prediction systems for two or more groups differ in any of the following ways: (1) predictive accuracy (standard errors of estimate); (2) the relationships (slopes) between the predictor and criterion; and (3) when the relationships are the same, the estimated criterion scores given the same predictor score (intercepts).

Purpose

This study examined the USNA selection composite to determine whether it is biased against Blacks.

Approach

The Black and White samples contained 350 and 5,155 midshipmen, respectively, from the USNA classes of 1986-1990. Two procedures were used sequentially to compare the Black and White prediction systems for potential race-related effects. The first procedure tests for differences in predictive accuracy (standard errors of estimate). The second procedure tests for differences in relationships (slopes) between the CM and academic performance and for differences in the prediction of academic performance scores given the same CM score (intercepts). The test for differences in intercepts should only be performed when differences in slopes are not statistically significant.

Results and Discussion

The predictive accuracy (standard errors of estimate) did not differ significantly for the Black and White samples. However, the test comparing the relationships (slopes) between the CM and academic performance did show a statistically significant difference for Blacks and Whites. While statistically significant, the minimal improvement in prediction that could be obtained by using separate prediction systems for Blacks and Whites would have no practical significance. Nevertheless, an effort was made to assess the impact of this statistically significant difference.

Toward this end, a regression line (a graphic representation of a prediction equation) was plotted for each racial group and the combined group. A comparison of these regression lines revealed that Black academic performance scores were generally overpredicted when the White

prediction equation was used to obtain these scores. The results were similar when the combined-group prediction equation was used to obtain Black academic performance scores. Therefore, use of the prediction equation that was developed on the combined sample of Black and White midshipmen does not disadvantage the Black minority.

Conclusion and Recommendation

For USNA applicants, separate selection composites for Blacks and Whites are not necessary. This conclusion is based on three findings.

1. The predictive accuracy of the CM is not significantly different for the two racial groups.
2. The difference in predictor-criterion relationships (slopes) between racial groups has no practical significance (i.e., the use of separate equations would not provide meaningful improvement in predicting academic performance).
3. Black academic performance scores are generally overpredicted when the combined-group prediction equation is used. Therefore, the use of a single selection composite does not adversely impact upon the Black minority.

It is therefore recommended that the USNA continue using a single prediction system.

CONTENTS

	Page
INTRODUCTION	1
Problem and Background.....	1
Purpose.....	2
APPROACH	2
Sample	2
Predictor.....	2
Criterion.....	2
Procedures.....	2
RESULTS AND DISCUSSION	3
Descriptive Statistics.....	3
Comparison of Standard Errors of Estimate	4
Comparison of Slopes.....	4
Comparison of Regression Lines.....	5
CONCLUSION AND RECOMMENDATION.....	7
REFERENCES	9

LIST OF TABLES

1. Means, Standard Deviations, and Correlations by Race for the USNA Classes of 1986-1990.....	4
2. Moderated Multiple Regression for Blacks and Whites in the USNA Classes of 1986-1990.....	5
3. Regression Equations for Blacks, Whites, and the Combined Group in the USNA Classes of 1986-1990.....	6

INTRODUCTION

Problem and Background

The general belief that tests are biased against minorities has been prevalent for many years. According to Linn and Hastings (1984), the results of most studies refute this belief. The evidence reveals either no difference in prediction for subgroups or overprediction of minority-group performance when it is predicted using majority-group equations (Linn, 1982; Schmidt & Hunter, 1981). Nevertheless, it is important to assess the fairness of the United States Naval Academy (USNA) selection process to ensure that this specific situation is consistent with the available evidence.

The USNA selection process uses a single selection composite, the Candidate Multiple (CM), to predict academic and other aspects of USNA performance for applicants, minority and majority. This procedure raises the question of whether it is appropriate to combine USNA midshipmen from different racial groups into a single sample when developing a system to predict academic performance for USNA applicants. An assessment of differential prediction addresses this question while clarifying some of the issues involved.

Differential prediction occurs when the prediction systems for two or more groups vary in terms of standard errors, slopes, and/or intercepts. Linn (1982, p. 374) described the implications of differences in standard errors of estimate, slopes, and intercepts in this way:

Differences in standard errors of estimate imply that the magnitude of the nonsystematic errors of prediction tends to be larger for members of one group than for members of another. . . . Differences in slope imply that the predicted criterion scores changed more with changes in the test score for the group with the steeper slope than they did for the group with the flatter slope. With equal slopes, a difference in intercepts implies that the predicted criterion score associated with any particular test score is higher for one group than the other.

Information regarding differences in prediction systems allows for informed consideration of fairness issues (Linn, 1978). In recent years, debates involving the fairness of selection models have abated. The regression model as defined by Cleary (1968) has become the most widely accepted of these models. Cleary (1968, p. 115) stated:

A test is biased for members of a subgroup of the population if, in the prediction of a criterion for which the test was designed, consistent nonzero errors of prediction are made for members of the subgroup. In other words, the test is biased if the criterion score predicted from the common regression line is consistently too high or too low for members of the subgroup.

Thus, according to Cleary, regression analyses provide a sound basis for the assessment of differential prediction.

Purpose

This study examined the USNA selection composite to determine whether it is biased against Blacks.

APPROACH

Sample

The Black and White samples contained 350 and 5,155 midshipmen, respectively, from the USNA classes of 1986-1990. These Black and White midshipmen declared themselves to be "Black" and "Caucasian/White," respectively, when applying for admission to the USNA. Blacks were chosen as the comparison minority group because they provided sufficient sample size for the statistical assessment of differential prediction. Other minority groups (i.e., Hispanics, Asians, Pacific Islanders, Native Alaskans, American Indians, Puerto Ricans, Filipinos, and those declaring themselves as "other") were excluded due to their small sample sizes.

Predictor

The key predictor of success at the USNA is the CM. The CM is an empirically-developed composite that weights Scholastic Aptitude Test/American College Test scores, high school class rank, recommendations, extracurricular activities, and vocational interest scales to predict academic performance, military performance, major, and disenrollment criteria (Alf, Neumann, & Mattson, 1988).

Criterion

An important measure of success at the USNA is the Academic Quality Point Ratio (AQPR). The AQPR is the measure of academic performance used in this report. It represents the academic cumulative grade point average earned by each midshipman after one year of study at the USNA. If a midshipman disenrolled before the end of the plebe year, the first semester AQPR was used when available.

Procedures

Two procedures were used sequentially to compare the Black and White prediction systems for potential race-related effects. The first procedure tests for standard-error-of-estimate differences. The standard error of estimate represents the standard deviation of prediction errors. These standard deviations for independent samples can be compared by employing a formula provided by McNemar (1969, p. 94, equation 6.11).

The second procedure, moderated multiple regression, tests for both slope and intercept differences. This procedure is similar to the one provided by Pedhazur (1982). The process uses a regression equation:

$$Y' = b_1X_1 + b_2X_2 + b_3X_1X_2 + A$$

where	Y'	=	predicted criterion score
	b_1	=	regression weight for main predictor
	X_1	=	main predictor
	b_2	=	regression weight for moderator variable
	X_2	=	moderator variable
	b_3	=	regression weight for product variable
	X_1X_2	=	product (interaction) of main predictor and moderator variable
	A	=	constant

The moderated multiple regression process includes three steps. First, the main predictor (X_1) is entered into the regression equation and r^2 is computed. Second, the moderator variable (X_2) is added to the equation. These two terms must be entered into the equation initially to determine the variance accounted for (R^2) before the interaction term (X_1X_2) can be assessed in the next step. Third, X_1X_2 is added to the regression equation to ascertain whether the interaction between the main predictor and the moderator is statistically significant. If the interaction (X_1X_2) is significant, then the groups have significantly different slopes, and it would be inappropriate to test for a difference in intercepts. If the interaction is not significant, it is then appropriate to assess whether the moderator (X_2) alone provides a significant increment in R^2 over the main predictor. A significant increment in R^2 would indicate significantly different intercepts.

For the present study, this three-step procedure involved the assessment of increments in R^2 that occurred when AQPR was predicted using three variables: the CM, a dichotomized moderator variable for race (RACE), and the interaction of the CM and RACE (CM X RACE).

To summarize, the comparison of prediction systems was done sequentially with standard errors of estimate assessed first. Slopes were assessed next if and only if the standard errors of estimate were not significantly different. Finally, the intercepts were assessed if and only if the other comparisons showed no significant differences (Reschly & Sabers, 1979).

RESULTS AND DISCUSSION

Descriptive Statistics

Means, standard deviations, and correlations for the CM and AQPR are presented by race in Table 1. The mean predictor and criterion scores for Blacks were about one standard deviation below the corresponding mean scores for Whites. While the validity coefficients were significant for both groups, the correlation between the CM and AQPR for Blacks was significantly lower ($p < .001$) than for Whites. However, differences in validity coefficients are not sufficient to establish differential predictability (e.g., Bartlett, Bobko, Mosier, & Hannan, 1978; Linn, 1978).

Table 1

**Means, Standard Deviations, and Correlations
By Race for the USNA Classes of 1986-1990**

	Predictor (CM)		Criterion (AQPR)		Validity (r)
	Mean	S.D.	Mean	S.D.	
Black (N = 350)	58,914	3,451	2.20	.45	.27*
White (N = 5,155)	63,212	4,237	2.64	.55	.46*
Total (N = 5,505)	62,938	4,320	2.61	.56	.47*

* $p < .01$

Comparison of Standard Errors of Estimate

The standard errors of estimate for the Black (.44) and White (.49) samples were not significantly different. As a result, it was appropriate to test for a difference in slopes.

Comparison of Slopes

The CM accounted for 22.29 percent of the variance in AQPR scores (see Table 2). RACE was added to the equation next and accounted for an additional 0.62 percent of the variance in AQPR scores. While this increment appears to indicate a statistically significant difference in intercepts, it is, as stated earlier, inappropriate to interpret this increment when slopes differ. When CM X RACE was added to test for a difference in slopes, it yielded a statistically significant increment in R^2 , accounting for an additional 0.14 percent of the variance in AQPR scores. While the negligible increment in R^2 was statistically significant, it is too small to be of practical significance. Nevertheless, an effort was made to assess the impact of this interaction in terms of predictive bias.

Table 2
Moderated Multiple Regression for Blacks and Whites
In the USNA Classes of 1986-1990

Predictor(s)	Criterion (AQPR)		
	Total R	Total R ²	Increment in R ²
CM	.4721**	.2229**	.2229**
CM, RACE	.4787**	.2291**	.0062**
CM, RACE, CM X RACE	.4801**	.2305**	.0014*

*p < .005

**p < .001

Notes. N for Blacks = 350.
 N for Whites = 5,155.

Comparison of Regression Lines

Having found a statistically significant interaction between the CM and RACE, separate regression equations were developed for each racial group and the combined group (see Table 3). These equations were used to plot the regression lines shown in Figure 1. It was then possible to determine the two regions along the CM continuum for which predicted AQPR values differed significantly for Blacks and Whites. These regions of significance were found by using the Johnson-Neyman technique (Johnson & Neyman, 1936). Observed CM scores ranged from 48,173 to 76,258 with scores greater than 54,889 predicting significantly higher AQPR values for Whites than for Blacks (see Figure 1). The other region of significance (i.e., CMs < 38,024) was determined to be far below the minimum observed CM (48,173) for the current samples. Therefore, the CM values in the lower region were not germane for the present study.

The regression lines in Figure 1 reveal that Black AQPR scores are significantly overpredicted by all CM values exceeding 54,889 when the White regression equation is used for the prediction of Black criterion scores. This overprediction occurs for 88 percent of the Blacks. When such overprediction occurs, members of the overpredicted subgroup obtain higher predicted criterion scores from another group's regression equation than would be obtained from their own group's regression equation. Similarly, the Black AQPR scores are overpredicted by the combined-group regression equation, but the Johnson-Neyman technique is not applicable when overlapping groups are compared. This result indicates that use of the combined-group regression equation does not disadvantage the Black minority. The White AQPR scores predicted from the combined-group regression equation are similar to the scores predicted from their own group's regression equation. This similarity is expected because of the 94 percent White representation in the combined group.

Table 3

Regression Equations for Blacks, Whites, and the Combined Group in the USNA Classes of 1986-1990

Group	Regression Equation			
Black	$Y' =$	$.0000351X$	$+$	$.129965$
White	$Y' =$	$.0000594X$	$-$	1.116423
Combined	$Y' =$	$.0000609X$	$-$	1.222959

Notes. Y' = predicted criterion score (AQPR).
 X = predictor score (CM).

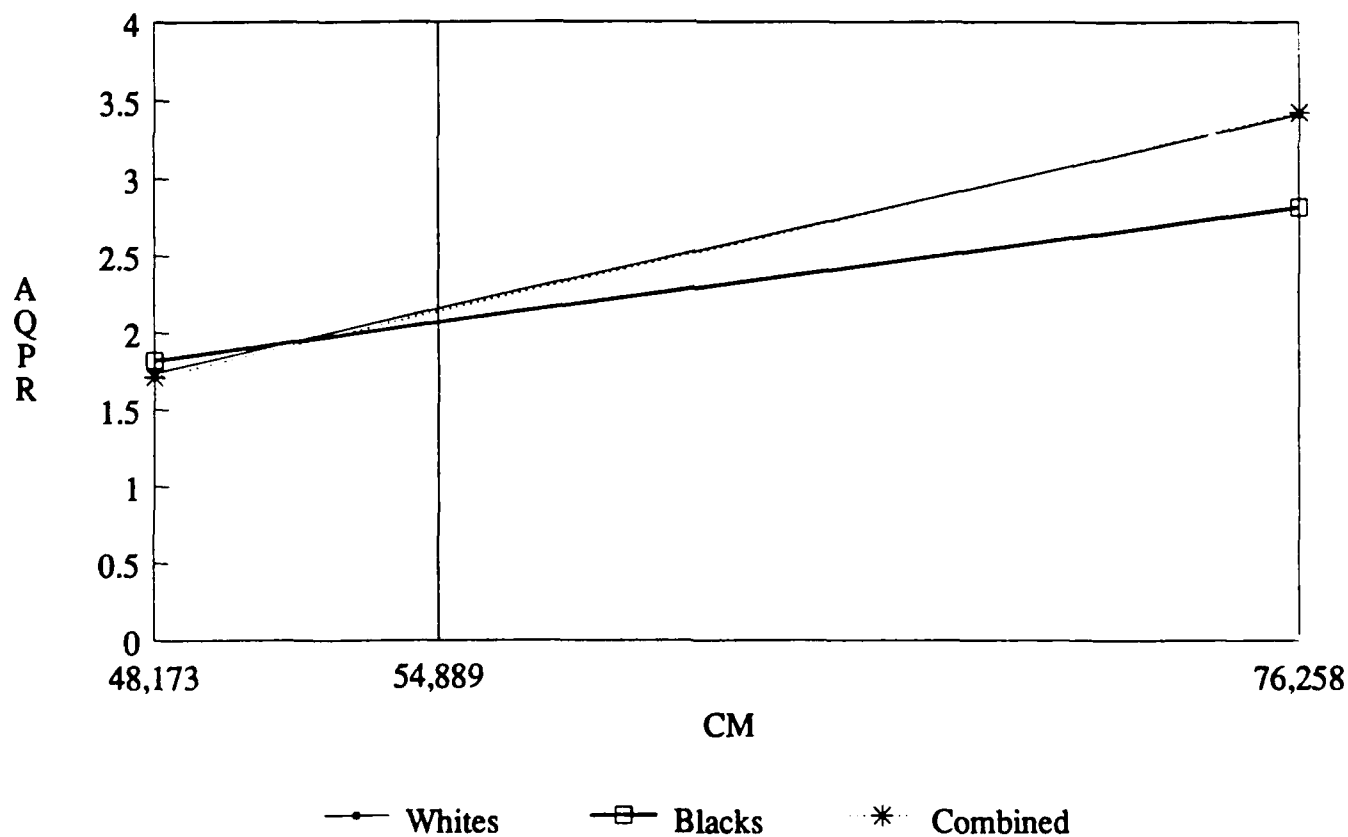


Figure 1. Regression lines to predict AQPR from CM.

CONCLUSION AND RECOMMENDATION

For USNA applicants, separate selection composites for Blacks and Whites are not necessary. This conclusion is based on three findings.

1. The predictive accuracy of the CM is not significantly different for the two racial groups.
2. The difference in CM-AQPR relationships (slopes) between racial groups has no practical significance (i.e., the use of separate equations would not provide meaningful improvement in predicting academic performance).
3. Black AQPR scores are generally overpredicted when the combined-group regression equation is used. Therefore, the use of a single selection composite does not adversely impact upon the Black minority.

It is therefore recommended that the USNA continue using a single prediction system.

REFERENCES

- Alf, E. F., Neumann, I., & Mattson, J. D. (1988). *Revision of the United States Naval Academy selection composite* (NPRDC Tech. Note 88-61). San Diego: Navy Personnel Research and Development Center.
- Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology*, 31, 233-241.
- Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their applications to some educational problems. *Statistical Research Memoirs*, 1, 57-93.
- Linn, R. L. (1978). Single-group validity, differential validity, and differential prediction. *Journal of Applied Psychology*, 63, 507-512.
- Linn, R. L. (1982). Ability testing: Individual differences and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies Part II* (pp. 335-388). Washington, DC: National Academy Press.
- Linn, R. L., & Hastings, C. N. (1984). Group differentiated prediction. *Applied Psychological Measurement*, 8, 165-172.
- McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: John Wiley & Sons.
- Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd ed.). Fort Worth, TX: Holt, Rinehart, and Winston.
- Reschly, D. J., & Sabers, D. L. (1979). Analysis of test bias in four groups with the regression definition. *Journal of Educational Measurement*, 16, 1-9.
- Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36, 1128-1137.

DISTRIBUTION LIST

Distribution:

Director, Office of Naval Technology (Code 20)

Office of Naval Technology (Code 222)

Commanding Officer, Naval Education and Training Program Management Support Activity
(Code 04) (2)

Defense Technical Information Center (DTIC) (2)

Copy to:

Assistant for Manpower Personnel and Training Research and Development (OP-01B2)

Director, Total Force Training and Education Policy Division (OP-114) (2)

Commander, U.S. ARI, Behavioral and Social Sciences, Alexandria, VA (PERI-POT-I) (2)